Can threat information bias fear learning? Some tentative results and methodological considerations

Gaëtan Mertens[a,b] and Jan De Houwer[a]

[a]Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

[b]Department of Clinical and Health Psychology, Utrecht University, Utrecht, the Netherlands

Correspondence concerning this article should be addressed to Gaëtan Mertens, Department of Clinical and Health Psychology, Heidelberglaan 1, room H1.29, Utrecht University, 3584CS Utrecht, the Netherlands.

E-mail: g.mertens@uu.nl

Tel: +31 30 253 75 53

Abstract

Whereas it is widely recognized that both verbal threat information and stimulus pairings can install strong and persistent fear, few studies have addressed the interaction between these two pathways of fear. According to the expectancy bias of Davey (1992, 1997), verbal information can install expectancy biases for aversive events that can result in facilitated fear learning through stimulus pairings and can delay extinction of fear. However, these predictions of the expectancy bias account have not been explored fully. Following up on two earlier studies (Field & Storksen-Coulson, 2007; Ugland, Dyson, & Field, 2013), we investigated the impact of prior threat information on fear acquisition, extinction and reinstatement. To this aim, participants received instructions about four unfamiliar animals, two of which that were described as dangerous whereas the other two were described as harmless. One animal of each pair was subsequently paired with an electric stimulus. Our results indicated that threat information resulted in stronger fear responses prior to fear conditioning and in delayed extinction of fear. However, these effects of instructions were not very pronounced and not found on all measures of fear. We discuss several methodological and procedural considerations that may modulate the effects of (verbally installed) expectancy biases.

Keywords: Instructions; Bias; Threat; Fear Conditioning; Extinction; Reinstatement; Expectancies; Startle Response; Skin Conductance Response

Can threat information bias fear learning? Some tentative results and methodological considerations

Both verbal threatening information, more specifically the verbal specification of the presence of potentially aversive or harmful stimuli in the presence of certain antecedent stimuli, and stimulus pairings, more specifically experiencing an aversive unconditioned stimulus (US) in the presence of a conditioned stimulus (CS), can result in strong and persistent fear responses (Field, 2006; Rachman, 1977). This claim is supported by retrospective studies with anxiety patients (King, Eleonora, & Ollendick, 1998; Merckelbach, de Jong, Muris, & van den Hout, 1996; Schindler, Vriends, Margraf, & Stieglitz, 2016) as well as laboratory studies demonstrating effects of both pathways on all three of Lang's (1968) fear response systems (Field, 2006; Muris & Field, 2010; Olsson & Phelps, 2004). However, most often these different pathways of fear acquisition are studied in isolation. Few studies investigated the interaction between them. This is problematic for our understanding of the etiology of fear because the effects of the different pathways of fear learning are unlikely to be completely isolated in daily life.

Several theories of fear learning predict that verbal information and stimulus pairings can have interactive effects (Davey, 1992, 1997; Field, 2006; Mineka & Zinbarg, 2006). Specifically, according to the expectancy model of Davey (1992, 1997), verbal instructions can install a bias to expect aversive events in the presence of certain antecedent stimuli. Such an expectancy bias can result in stronger autonomic fear reactions to these stimuli prior to any stimulus pairings. Furthermore, Davey (1992) argued that continued non-reinforcement can eliminate this expectancy bias, while threat or actual administration of an aversive event (i.e., US) maintains this bias. Davey's expectancy bias theory was initially proposed to account for biased fear learning observed for evolutionary fear-relevant stimuli (e.g., pictures of snakes and spiders;

Öhman & Mineka, 2001). Davey argued that expectancy biases, rather than selective associations (Öhman & Mineka, 2001; Seligman, 1971), drive the biased learning for fear-relevant stimuli and that these expectancy biases observed for evolutionary fear-relevant stimuli do not necessarily need to be the result of evolutionary selection, but can also be due to cultural learning, of which instructions are an intrinsic component (Tomasello, Kruger, & Ratner, 1993). Based on this expectancy bias model of Davey (1992, 1997) it can be predicted that verbal instructions can install expectancy biases which in turn can result in prepared-like learning effects such as those that have been observed for so-called fear-relevant stimuli. Specifically, the expectancy bias model predicts that verbal threatening instructions result in facilitated learning through stimulus pairings and resistance to fear extinction when USs are no longer administered.

So far, to our knowledge, only two studies have explicitly addressed these predictions. In a first study by Field and Storksen-Coulson (2007), young children (six to eight years old) were given negative information about an unknown animal. In a subsequent task, the children either had a negative encounter with that animal (a sudden movement in a box where the animal was told to be in), or with an animal they did not receive information about. The results indicated that children displayed more avoidance when threat information and a negative encounter were combined compared to the isolated effects of threat information or conditioning experience. In another study by Ugland, Dyson and Field (2013), participants received negative information about two animals. Furthermore, two animals were included in the study about which they did not receive any information. In a subsequent acquisition phase, one of the two threatened animals and one of the two neutral animals were paired with an unpleasant sound, while the other two (threatened and neutral) animals were paired with the unpleasant sound or a pleasant sound in a 50:50 ratio. Expectancy of the unpleasant sound, fear beliefs about the animals and event-related

potentials (ERP) were collected as measures of fear in their study. Ugland et al. (2013) found that the participants more quickly (i.e., in less trials) learned to expect the unpleasant sound for the threatened animal than for the neutral animal. However, on the ERP components and on the fear beliefs measure no such interaction between verbal instructions and negative experience was obtained. Furthermore, no evidence for an impact of threatening information on extinction of fear as measured by the ERP components and fear beliefs was found (expectancy ratings were not collected during the extinction phase in their study). Taken together, the results of the studies by Field and Storksen-Coulson (2007) and Ugland et al. (2013) partly support the expectancy model of Davey (1997) by demonstrating that combining threatening instructions with (mild) aversive stimulus pairings results in more avoidance (Field & Storksen-Coulson, 2007) and more quickly learned expectancy of an US (Ugland et al., 2013). However, so far this interaction between threat instructions and conditioning has not been successfully demonstrated on psychophysiological measures of fear and no evidence has been obtained that threat instructions can result in delayed extinction of fear. Nevertheless, these are important lacunas to address because physiological reactions are considered to be an independent and important aspect of fear (Lang, 1968), and because resistance to extinction is often thought of to reflect phylogenetic evolved hardwired associations, independent of language and higher order cognitions (Öhman & Mineka, 2001).

Our study served as a conceptual replication and extension of the studies of Field and Storksen-Coulson (2007) and Ugland et al. (2013). The general procedure of our experiment was similar to the study of Ugland et al. (2013), namely that participants were first given threatening information about unknown animals and control information about two other animals, and were then subjected to a conditioning procedure followed by an extinction phase. However, in our

own study we measured more common psychophysiological measures of fear (skin conductance responses and fear potentiated startle). Furthermore, expectancy ratings were included in the extinction phase to assess expectancy biases during all the phases of the experiment (acquisition, extinction and reinstatement). As such, our study allowed us to investigate whether interaction effects between verbal instructions and aversive experience can be obtained on psychophysiological measures of fear and whether verbal instructions can result in resistance to extinction.

Finally, we investigated another factor that was not investigated in any of the previously mentioned studies on the interaction between different learning pathways. More specifically, we investigated whether the degree of ambiguity with regard to the likelihood of aversive events may modulate the effects of prior threat instructions. Cognitive biases (such as expectancy biases) usually have their largest impact in ambiguous or uncertain situations (Chan & Lovibond, 1996; Lissek, Pine, & Grillon, 2006; Ly & Roelofs, 2009; Tversky & Kahneman, 1974). Accordingly, the effects of expectancy biases installed via instructions might be more pronounced when the ambiguity of the situation is high relative to when it is low. In order to operationalize ambiguity, we manipulated the reinforcement rate in the acquisition phase of the experiment. Half of the participants received an electric stimulus each time when they were presented a conditioned stimulus (CS+) during the acquisition phase (continuous reinforcement schedule) whereas the other half of participants only received an electric stimulus half of the time when they were presented a CS+ (partial reinforcement schedule). We expected that any effects of verbal threat instructions such as facilitated learning or delayed extinction would be most pronounced with the condition with the partial reinforcement schedule because there is

more ambiguity with regard to the presence of the aversive event in this condition compared to the condition with the continuous reinforcement schedule.

Method

Participants

Seventy-seven right-handed students (19 men, 58 women) at Ghent University participated in exchange for €10. From this initial sample, the data of 15 participants was excluded because they indicated that they did not believe the threat instructions and the data of one participant was excluded because the electrode for administering the electric stimulus became detached (see the *Data reduction and analysis* section). Demographic details for the final sample in each of the conditions of this experiment is provided in Table 1. All participants completed an informed consent form and were instructed that they could discontinue the experiment at any point without any negative consequences. This study was approved by the ethics committee of the Faculty of Psychology and Educational Sciences at Ghent University.

Table 1. Demographic information for the final sample of participants in each of the conditions of the experiment. Standard deviations between brackets.

	Continuous reinforcement (N = 29)	Partial Reinforcement (N = 33)	Difference
Mean age	24.97 (6.32)	23.31 (3.34)	$t(60) = 1.32$
Sex distribution	6 males	8 males	$X^2(1) < 1$
Mean STAI-T	36.55 (10.03)	36.72 (7.78)	$t(60) < 1$
Mean electric stimulus intensity (mA)	5.56 (2.24)	6.15 (2.29)	$t(60) = 1.02$
Mean electric stimulus pain rating	7.90 (0.65)	7.97 (0.70)	$t(60) < 1$

Material

Conditioned Stimuli. CSs were four pictures (520 by 390 pixels) of animals (pizote, cuscus, quoll and aye-aye; see Ugland, Dyson, & Field, 2013) presented in the middle of a 17 inch Dell computer screen (resolution: 1024 by 768 pixels). Assignment of these animal pictures to the different CS types (threatened CS+, threatened CS-, neutral CS+, neutral CS-) was randomized over participants.

Unconditioned Stimulus. The US was an electric stimulus that consisted of 10 rectangular pulses of 2 ms with and inter pulse interval of 8 ms, creating an electric stimulus of 100 ms. This stimulus was administered by two lubricated Fukuda standard Ag/AgCl electrodes (1-cm diameter; inter-electrode distance: ~2-cm) to the left leg over the retromalleolar course of the

sural nerve. The stimulus was generated by a constant current stimulator (DS7A, Digitimer, Hertfordshire, UK). The intensity of the electric stimulus was determined for each participant individually to be unpleasant but not painful using a stepwise work-up procedure (see the Procedure section for details concerning this work-up procedure).

Psychophysiology

Skin Conductance Responses (SCRs). SCRs were collected using a Coulbourn V71-23 skin conductance coupler (Coulbourn Instruments, Allentown, PA) and disposable Ag/AgCl electrodes (3M Red Dot 2259-50, 17 mm diameter) attached to the thenar and hypothenar eminences of the non-dominant hand. The signal was measured using a constant voltage coupler (0.5 V DC coupling) and digitized at 10 Hz. The collected data were smoothed and further analyzed offline with Psychophysiological Analysis (PSPHA) (De Clercq, Verschuere, De Vlieger, & Crombez, 2006). SCRs were calculated by subtracting the lowest amplitude that occurred prior to the highest amplitude (0 to 7 seconds interval) from the highest amplitude within a 1 to 7 seconds interval after CS onset. This deviation from the standard 2-seconds baseline entire-interval scoring method (e.g., Milad, Orr, Pitman, & Rauch, 2005; Pineles, Orr, & Orr, 2009; Raes, De Houwer, De Schryver, Brass, & Kalisch, 2014; Soeter & Kindt, 2012) was preferred due to the high baseline values in this study, especially during the extinction phase, which produced an overly conservative scoring method using the standard entire-interval scoring method (i.e., a very high number of zero-responses). Values smaller than 0.02μS were recoded to zero. Finally, collected SCRs were range corrected with the highest recorded amplitude for that participant to account for individual differences in responsivity (Lykken & Venables, 1971) and square root transformed to normalize the data (Dawson, Schell, Filion, & Berntson, 2007).

Fear Potentiated Startle (FPS). FPS was measured using two miniature Ag/AgCl electrodes (0.5 cm diameter) filled with conductive gel. One electrode was placed just below the pupil of the left eye and the other electrode was placed approximately 1 cm laterally. A ground electrode was placed in the middle of the forehead (Blumenthal et al., 2005). Electrode sites were first gently cleaned with scrub gel and water. The raw electromyographic signal was amplified 50,000 times, filtered online (band pass: 13 – 1000 Hz) and digitally stored at 1000 Hz using a Coulbourn V75-01 bioamplifier (Coulbourn Instruments, Allentown, PA). The acquired data were rectified and smoothed in the area of interest (0 – 150 ms after probe onset) with a FIR filter (Nitschke, Miller, & Cook, 1998) using PSPHA. The startle probe was a 50 ms white noise burst (104 dB) generated using a V85-05C Coulbourn audio module and administered via Sennheiser headphones.

The acquired signal was scored semi-automatically using PSPHA by subtracting the mean baseline value (0 - 20 ms after probe onset) from the peak value in the 21 - 150 ms window after probe onset. All startle responses were visually inspected and scored as missing values if a voluntary blink occurred just before, during or after probe onset, or if there were any other artifacts obscuring the startle response. The scores were subsequently T-transformed to control for inter-individual differences in responsitivity.

Questionnaire

The trait version of the State-Trait Anxiety Inventory (STAI-T; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983; van der Ploeg, Defares, & Spielberger, 2000) was used to determine the general anxiety level of the participants.

Procedure

Work-up procedure. After filling in the informed consent form and the STAI-T questionnaire, participants first went through a work-up procedure to determine the intensity level of the electric stimulus. During this procedure, participants were exposed to gradually increasing stimulus intensity levels and were asked to report on their experience. Specifically, participants were asked after each intensity level to verbally rate the electric stimulus on a painfulness scale ranging from zero (not painful at all) to ten (maximally tolerable pain). A minimal painfulness threshold for the electric stimulus was set at seven. The procedure was stopped when participants indicated that they felt uncomfortable experiencing higher intensities of the electric stimulus. If a participant gave a rating of less than seven and indicated that he or she did not want to experience a more intense electric stimulus, the work-up was also stopped and the stimulus with the highest tolerated intensity was selected (see Mertens & De Houwer, 2016 for prior use of this work-up procedure). The final selected electric stimulus intensity levels ranged between 2.2 and 22 mA (M = 6.10, SD = 2.91) and pain ratings ranged between 4 and 9.5 (M = 7.86, SD = 0.89) (see Table 1 for the corresponding electric stimulus intensity levels and pain ratings for the final sample of participants in each condition of the experiment). After the work-up procedure, psychophysiology recording electrodes were applied as described above. Finally, headphones for the startle probe administration were put on. Participants were verbally informed that these headphones served to present loud but harmless noises to them throughout the experiment.

Threat instructions and test. At the start of the experiment participants were asked to read the descriptions of a number of animals. They were instructed to memorize this information

because they would be tested about it. Subsequently, participants saw two pictures of animals, one in the right and one in the left upper corner of the screen, together with threatening information about these animals. The information stated that these animals were aggressive towards humans and pets and could potentially carry diseases. On the next instruction page, participants again saw two animals in the top right and left corners of the screen together with neutral information about these animals such as where they live, what they eat and what their behavior is (see the appendix for the full translation of the threatening and neutral information). Each time, participants could move through these descriptions at their own pace by pressing the spacebar.

After this information, participants were required to complete a short test. The test consisted of two trials in which two pictures of the animals were presented on the left and right hand side of the screen. Participants were required on each trial to select the animal that was potentially dangerous for humans. After clicking one of the two pictures they received feedback for 400 ms indicating whether they had selected the right option. If participants made a mistake on either of the two trials, they received the instructions and the test again until they passed it (mean number of tests until pass = 1.05, SD = 0.28, Range = 1-3).

Acquisition, extinction and reinstatement phase. When participants successfully completed the test they moved on to the acquisition, extinction and reinstatement phases of the experiment. At the start of the acquisition phase participants were told that they would see pictures of animals and that some of these animals could be followed by an electric stimulus, while other animals would not. They were instructed to attentively follow which animals were followed by the electric stimulus and which were not followed by the electric stimulus.

Furthermore, they were instructed to indicate to what extent they expected to feel an electric stimulus each time they saw a picture of an animal by clicking one of the options on the scale below the pictures.

After these instructions, the acquisition phase started. This phase started with six non-cued startle probes (ITI: seven seconds) to habituate participants to the probe. The acquisition phase consisted of 24 trials (six presentations of each of the four CSs). Each trial consisted of the presentation of a fixation cross during one second, followed by the CS during eight seconds. Startle probes were presented seven seconds after CS onset on each trial. During CS presentations a 9-point Likert scale (anchors: 1 = "not at all", 5 = "uncertain", 9 = "certainly"; caption = "To what extent do you expect the electric stimulus?") was presented below the CS which participants could use to indicate their expectancy of an electric stimulus (using a computer mouse and their dominant hand)[1]. Trial order was semi-randomized to limit the maximum number of consecutive identical trials to four. For one half the participants each CS+ trial was followed by the electric stimulus at CS offset (100% reinforcement rate), whereas for the other half of the participants each CS+ trial was followed by the electric stimulus at CS offset on half of the trials (50% reinforcement).

The extinction phase followed the acquisition phase without interruption. The extinction phase was identical to the acquisition phase with regard to the number of trials and trial order organization, however, no electric stimuli were administered during this phase.

---

[1] No response deadline was imposed for the US expectancy ratings, but participants were encouraged to respond as fast and as spontaneously as possible. One and a half percent of all US expectancy ratings were provided after CS offset (i.e., after more than 8 seconds). Inclusion of these responses in the data did not change the results and therefore we report the results including these delayed US expectancy ratings. One important reason for including these ratings in the final analyses is because repeated measures ANOVA's in SPSS do not take into account entries (i.e., participants) with missing data.

Finally, a reinstatement US was administered 13, 15 or 17 seconds after the last extinction trial. This reinstatement US was followed after 7.5 seconds by two more presentations of each CS (eight trials) in a random order without reinforcement.

Fear and believability ratings. At the end of the experiment, participants were asked to complete a few questions about how they felt when seeing the different animals used in the experiment. Each trial of this rating session consisted of a CS being presented in the middle of the screen with a 9-point Likert scale below it (anchors: 1 = not fearful at all, 5 = unsure, 9 = very fearful; caption = "How fearful were you while seeing these animals?"). A fear rating was collected for each CS and trial order was randomized.

Finally, participants were asked to indicate how believable they found the information about the animals from a dropdown list of options ('not believable', 'not very believable', 'very believable', 'fully believable') and to indicate whether they were familiar with one or several of the animals ('yes', 'no', 'unsure'). Finally, participants had the option to write down anything they might not found believable in a text box before concluding the experiment.

Data reduction and analysis

The data of one participant was excluded because the electric stimulus electrode became detached during the experiment. Furthermore, the data of 15 participants were excluded because they reported insufficient belief of the threat information (see below)[2]. Finally, the startle data of an additional six participants and the SCR data of 12 participants were excluded for these

[2] Exclusion of these participants did not affect the conclusions of this experiment. Where relevant, the results including these subjects will be mentioned in footnotes.

measures only because the data quality was judged to be insufficient (i.e., more than 85% missing or zero responses for SCRs and more than 50% for startle responses)[3].

All collected fear measures (fear ratings, expectancy ratings, FPS, SCR) were analyzed using repeated factors ANOVA's with threat instructions (threatened, neutral) and CS type (CS+, CS-) as within-subject factors, and schedule (continuous reinforcement, partial reinforcement) as a between-subject factor. Additionally, for the fear measures that were measured continuously (expectancy ratings, SCR, FPS) the data were analyzed separately for each phase of the experiment (acquisition, extinction, reinstatement) with an additional factor trial (for expectancy ratings: trial 1 to 6; for SCR and FPS: trial 1 to 3; for the reinstatement analysis: last trial before the reinstatement US versus the first trial after the reinstatement US). The data of the psychophysiological measures was averaged per two trials to minimize the impact of missing data points. Greenhouse-Geisser corrections are reported when the sphericity assumption was violated.

Finally, we investigated the impact of trait anxiety on our results. Elevated trait anxiety scores have been related to different indexes of fear acquisition and extinction such as increased fear responses towards the CS+ and CS- (compared to inter-trial interval fear responses), increased differential conditioning (Duits et al., 2015; Lissek et al., 2005), delayed extinction of fear (Liberman, Lipp, Spence, & March, 2006) and increased return of fear (Kindt & Soeter, 2014). However, these patterns are often weak and inconsistent over studies (e.g., Kuhn, Mertens, & Lonsdorf, 2016; Torrents-Rodas et al., 2013). In our study we investigated the impact

[3] Note that we applied a more lenient criterion for data quality for SCRs than for FPS because zero responses are much more common for SCRs (Dawson et al., 2007) than for startle reflexes.

of trait anxiety on differential fear acquisition (fear responses for CS+ minus fear responses for CS- for the last two trials of the acquisition phase, providing an asymptotic index of fear acquisition), fear extinction (fear responses for CS+ minus fear responses for CS- for the last two trials of the extinction phase, providing an asymptotic index of fear extinction), return of fear for the CS- (fear responses for CS- after reinstatement minus fear responses for CS- prior reinstatement, providing an index for the return of fear for the CS- while controlling for pre-reinstatement fear for the CS-) and return of fear for the CS+ (fear responses for CS+ after reinstatement minus fear responses for CS+ prior reinstatement, providing an index for the return of fear for the CS+ while controlling for pre-reinstatement fear for the CS+). Furthermore, we investigated whether these relations between trait anxiety and these indexes of fear acquisition and extinction depended on the used reinforcement rate (continuous versus partial reinforcement) and on prior threat information (threatened versus neutral CSs) by including mean-centered trait anxiety scores as a covariate in a repeated measures ANCOVA with these two factors as a between and a within-subjects factor, respectively. Because of brevity and because this was not the main interest of this study, the results of these analyses are not discussed in the main text, but are included in the supplementary material.

Results

Believability of the information and familiarity with the animals

Most participants reported the threat information to be either ‘very believable’ (52 participants, 67.53%) or ‘fully believable’ (10 participants, 12.99%). However, some participants also reported to find the threat information ‘not very believable’ (13 participants, 16.88%) or ‘not believable’ (2 participants, 2.60%). The data of participant who indicated the threat instructions

were ‘not very believable’ or ‘not believable’ were excluded from the analyses. Furthermore, most participants indicated that they were unfamiliar with the animals (66 participants, 85.71%). However, some participants indicated that they were familiar with the animals (10 participants, 12.99%) and one participant (1.30%) indicated that he or she was unsure whether he or she was familiar with the animals. The data of participants who indicated to be familiar with the animals or were unsure, but who did believe the instructions, were retained for the analyses (6 participants).

Fear ratings

Analyses of the fear ratings (which were collected only at the end of the experiment) revealed significant main effects of CS type, $F(1, 59) = 76.41$, $p < .001$, $\eta^2_p = .56$, and of threat information, $F(1, 59) = 5.43$, $p = .023$, $\eta^2_p = .08$. These results demonstrate that participants reported more fear for animals that were paired with the US and for animals about which they had received threatening information (see Figure 1). The only significant interaction effect was between CS type and schedule, $F(1, 59) = 6.39$, $p = .014$, $\eta^2_p = .10$. As can be seen from Figure 1, this interaction was due to a larger difference between CS+ and CS- in the condition with the continuous reinforcement schedule than in the condition with the partial reinforcement schedule. None of the other interaction or main effects were significant, F-values < 2.

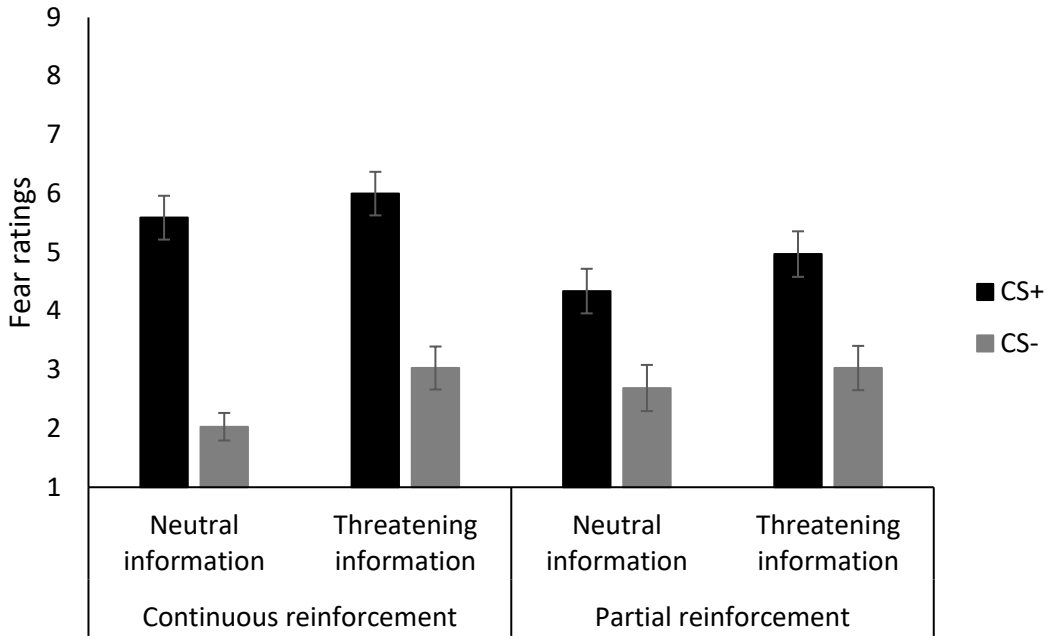

Figure 1. Fear ratings for the different CS types in the two conditions of the experiment.

US expectancy ratings

Acquisition phase. The analyses of the US expectancy ratings during the acquisition revealed a significant main effect of factors CS type, $F(1, 59) = 148.81$, $p < .001$, $\eta^2_p = .72$, threat information, $F(1, 59) = 19.41$, $p < .001$, $\eta^2_p = .25$, and trial, $F(5, 295) = 3.89$, $p = .002$, $\eta^2_p = .06$. Furthermore, significant two-way interaction effects were observed between CS type and schedule, $F(1, 59) = 37.14$, $p < .001$, $\eta^2_p = .39$, CS type and trial, $F(4.07, 240.32) = 46.57$, $p < .001$, $\eta^2_p = .44$, and threat information and trial, $F(3.93, 231.95) = 8.94$, $p < .001$, $\eta^2_p = .13$. Finally, a three-way interaction was observed between CS type, trial and schedule, $F(4.07, 240.32) = 18.62$, $p < .001$, $\eta^2_p = .24$. This latter interaction qualified most of the significant main and interaction effects and indicates that differences in US expectancy ratings between CS+ and CS- (i.e., higher US expectancy ratings for CS+ than for CS-) became more pronounced over

trials, especially in the condition with the continuous reinforcement schedule (see Figure 2). The interaction between threat information and trial was due to a slight decrease in US expectancy ratings for the threatened CSs over trials, whereas US expectancy ratings for the neutral CSs slightly increased over trials. In contrast to our hypotheses, the interaction effects between CS type and threat information, CS type, threat information and schedule, and CS type, threat information, trial and schedule were not significant, *F*-values < 1.7.

Extinction phase. Analyses of the US expectancy ratings during the extinction phase revealed significant main effects of CS type, $F(1, 59) = 61.99$, $p < .001$, $\eta^2_p = .51$, and trial, $F(2.56, 151.19) = 82.38$, $p < .001$, $\eta^2_p = .58$, and a marginally significant effect of schedule, $F(1, 59) = 3.58$, $p = .063$, $\eta^2_p = .06$. Furthermore, significant two-way interactions were observed between CS type and trial, $F(3.09, 182.01) = 20.37$, $p < .001$, $\eta^2_p = .26$, and between trial and schedule, $F(2.56, 151.19) = 11.79$, $p < .001$, $\eta^2_p = .17$. These two-way interactions were qualified by a three-way interaction between CS type, trial and schedule, $F(5, 295) = 10.03$, $p < .001$, $\eta^2_p = .15$, indicating that the reduction of differential US expectancy ratings between the CS+ and CS- over trials (i.e., extinction) was more pronounced in the condition with the continuous reinforcement schedule than in the condition with the partial reinforcement schedule (this is known as the partial reinforcement extinction effect). Finally, a marginally significant four-way interaction effect was found between CS type, trial, threat information and schedule, $F(3.88, 228.99) = 2.36$, $p = .056$, $\eta^2_p = .04$[4]. Explorative analyses revealed that this interaction was due to higher US expectancy ratings for the threatened CS- during the first two trials of the extinction phase in the condition with the partial reinforcement schedule, while US expectancy ratings for

[4] This interaction effect was significant when the subjects that did not find the threat information believable were included in the analyses, $F(4.07, 305.27) = 2.64$, $p = .033$, $\eta^2_p = .03$.

all other CSs during all trials in both reinforcement schedule conditions were highly similar (see Figure 2).

Reinstatement. Analyses of the US expectancy ratings just before and after the reinstatement manipulation revealed a significant effect of CS, $F(1, 59) = 33.29$, $p < .001$, $\eta^2_p = .36$, and trial, $F(1, 59) = 38.06$, $p < .001$, $\eta^2_p = .39$. Furthermore, the two-way interaction between schedule and trial, $F(1, 59) = 8.58$, $p = .005$, $\eta^2_p = .13$, was also significant. Finally, the three-way interaction effect between CS type, trial and schedule approached significance, $F(1, 59) = 3.68$, $p = .060$, $\eta^2_p = .06$. This pattern of results demonstrate that the reinstatement manipulation resulted in higher US expectancy ratings, especially in the condition with the continuous reinforcement schedule and was more pronounced for the CS+ in this condition (see Figure 2). The crucial interactions between CS type and threat information, CS type, schedule and threat information, and CS type, schedule, trial and threat information were not significant, F-values < 1.

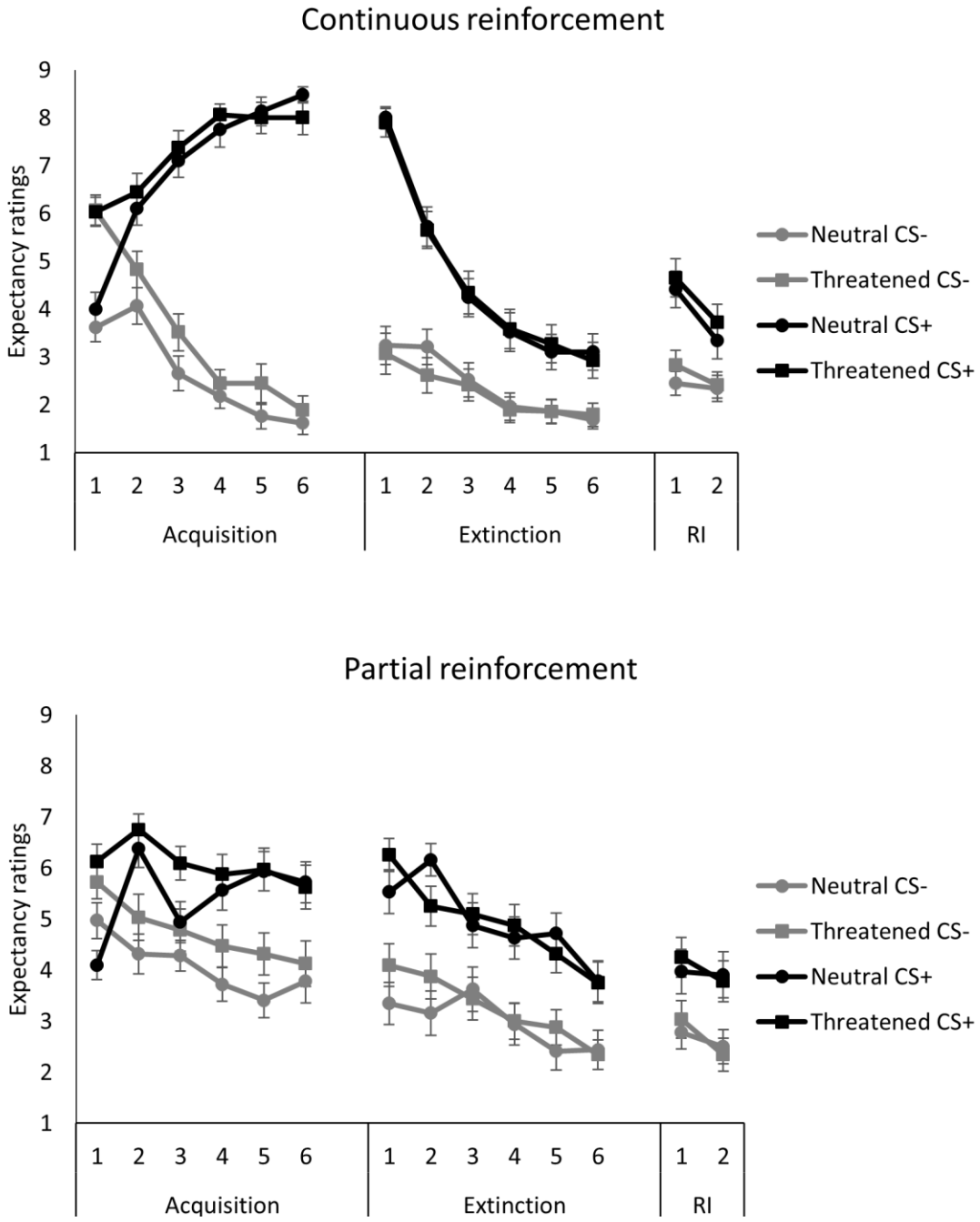

Figure 2. US expectancy ratings throughout the different phases of the experiment. RI = reinstatement.

SCRs

Acquisition phase. Analyses of the SCRs during the acquisition phase revealed significant main effects of trial, $F(2, 94) = 20.21$, $p < .001$, $\eta^2_p = .30$, and CS type, $F(1, 47) = 9.74$, $p = .003$, $\eta^2_p = .17$. These main effects demonstrate the habituation of the SCRs and the effect of the conditioning procedure (i.e., larger SCRs for CS+ than for CS-). Furthermore, significant two-way interactions were observed between CS and schedule, $F(1, 47) = 11.83$, $p = .001$, $\eta^2_p = .20$, and between threat information and schedule, $F(1, 47) = 4.75$, $p = .034$, $\eta^2_p = .09$. These interactions were due to larger differential SCRs between CS+ and CS- in the condition with the continuous reinforcement schedule and to larger SCRs for the neutral CSs in the condition with the continuous reinforcement schedule while SCRs were slightly larger for the threatened CSs in the condition with the partial reinforcement schedule (see Figure 3). None of the other main or interaction effects were significant, F-values < 1.2.

Extinction phase. None of the main or interaction effects were significant in the analyses of the SCRs during the extinction phase. The only potential meaningful data patterns were the main effects of threat information, $F(1, 47) = 2.80$, $p = .101$, $\eta^2_p = .06$, and of trial, $F(1.60, 75.17) = 2.28$, $p = .120$, $\eta^2_p = .05$. The pattern for threat information was in the opposite direction than what was hypothesized (i.e., slightly larger SCRs for neutral CSs compared to threatened CSs). All other F-values < 1.

Reinstatement. The analysis of the SCRs before and after the reinstatement manipulation revealed only a marginal significant effect of threat information, $F(1, 47) = 3.70$, $p = .061$, $\eta^2_p = .07$. As in the extinction phase, this effect was due to larger SCRs to neutral compared to threatened CSs. The other main and interaction effects were not significant, F-values < 1.9.

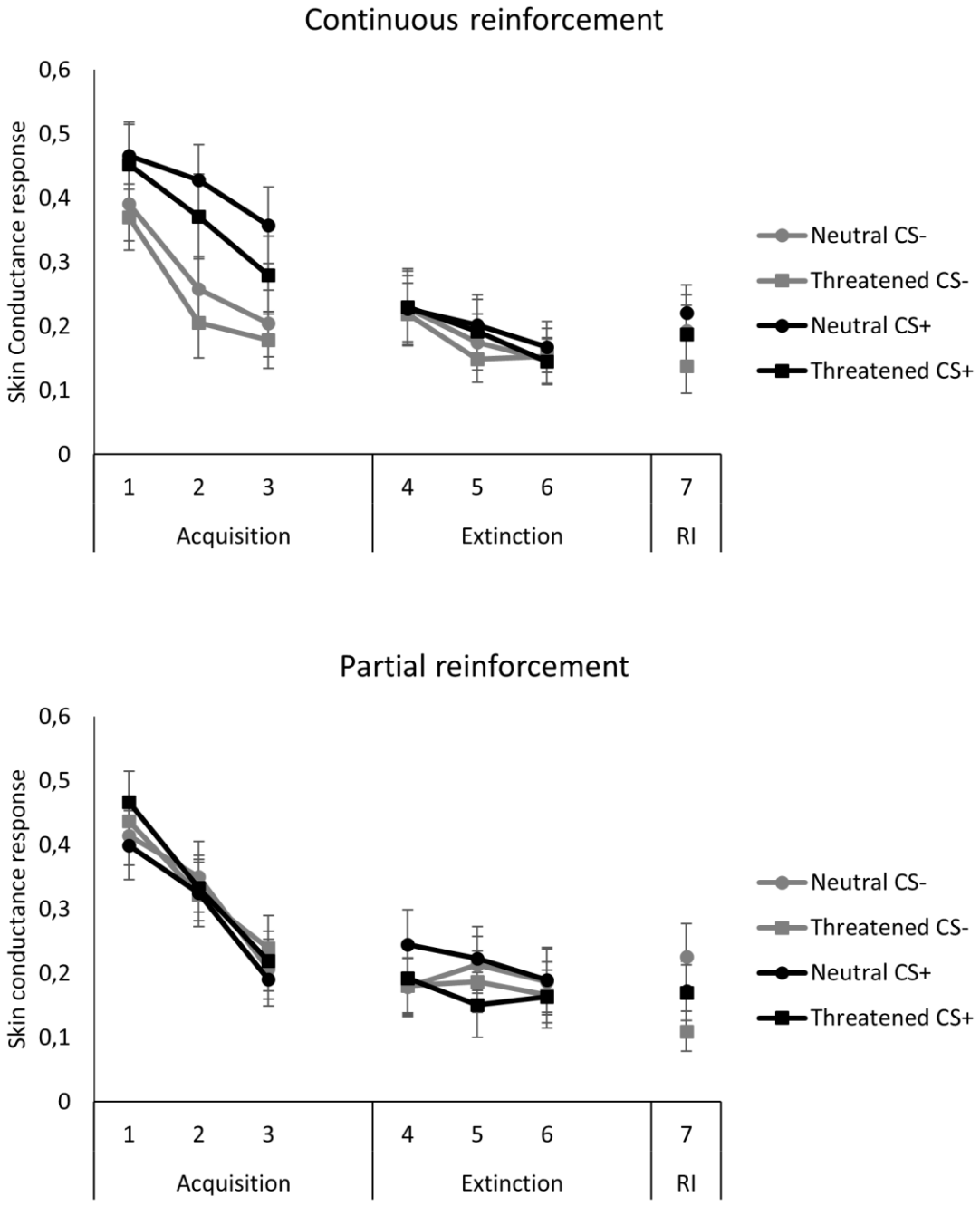

Figure 3. Range corrected and square root transformed skin conductance responses (averaged per two trials) throughout the different phases of the experiment. RI = reinstatement.

FPS

Acquisition phase. The analyses of the startle responses during the acquisition phase revealed a large effect of trial, $F(1.74, 92.41) = 48.02$, $p < .001$, $\eta^2_p = .48$, and a marginal significant effect of CS type, $F(1, 53) = 3.24$, $p = .077$, $\eta^2_p = .065$. The main effect of threat information did not reach significance, $F(1, 53) = 2.70$, $p = .107$, $\eta^2_p = .05$. These main effects illustrate a large habituation of the startle response during the acquisition phase, and a trend for an effect of the conditioning procedure (i.e., larger startle responses to CS+ than to CS-). Furthermore, significant two-way interactions were obtained between threat information and schedule, $F(1, 53) = 4.93$, $p = .031$, $\eta^2_p = .09$, and between CS type and trial, $F(2, 106) = 3.85$, $p = .024$, $\eta^2_p = .07$. These interactions indicate that the effect of threat information (i.e., larger startle responses for the threatened compared to the neutral CSs) was more pronounced in the condition with the continuous reinforcement schedule than in the condition with the partial reinforcement schedule and that the differential startle responses between CS+ and CS- at the start of the acquisition phase disappeared by the end of this phase (see Figure 4). Finally, a marginally significant three-way interaction was observed between threat information, trial and schedule, $F(1.82, 96.58) = 2.65$, $p = .081$, $\eta^2_p = .05$. This interaction was due to a larger habituation of the startle response for the threatened CSs over trials, especially in the condition with the continuous reinforcement schedule (see Figure 4). None of the other main or interaction effects were significant, F-values < 1.

[5] This main effect was significant when the subjects that did not find the threat information believable were included, $F(1, 65) = 4.00$, $p = .050$, $\eta^2_p = .06$.

Extinction phase. In this phase significant effects of trial, $F(2, 108) = 15.50$, $p < .001$, $\eta^2_p = .22$, and of CS type, $F(1, 54) = 4.87$, $p = .032$, $\eta^2_p = .08$, were observed. Again, these main effect illustrate the habituation of the startle response and the effect of the conditioning procedure (larger startle responses for CS+ than for CS-, see Figure 4). Furthermore, a marginally significant three-way interaction between threat information, trial and schedule, $F(1.77, 95.68) = 2.83$, $p = .070$, $\eta^2_p = .05$, and a significant four-way interaction between CS type, threat information, trial and schedule, $F(2, 108) = 4.06$, $p = .020$, $\eta^2_p = .076$, were observed. The latter interaction effect was due to the fact that differential startle responses (i.e., larger for CS+ than for CS-) were primarily observed at the start and middle of the extinction phase for the neutral CSs, whereas differential startle responses were obtained at the end of the extinction phase for the threatened CSs, especially in the condition with the partial reinforcement schedule (see Figure 4). The other main or interaction effects were not significant, F-values < 2.3.

Reinstatement. The only significant effect obtained in this analysis was the interaction between CS type and schedule, $F(1, 53) = 5.75$, $p = .020$, $\eta^2_p = .10$. This interaction was due to larger startle responses for the CS+s compared to the CS-s in the condition with the partial reinforcement schedule, whereas this pattern was reversed in the condition with the continuous reinforcement schedule (see Figure 4). None of the other main or interaction effects reached significance, F-values < 1.5.

---

[6] This interaction effect remained significant when the subjects that did not find the threat information believable were included, $F(2, 132) = 3.77$, $p = .026$, $\eta^2_p = .05$

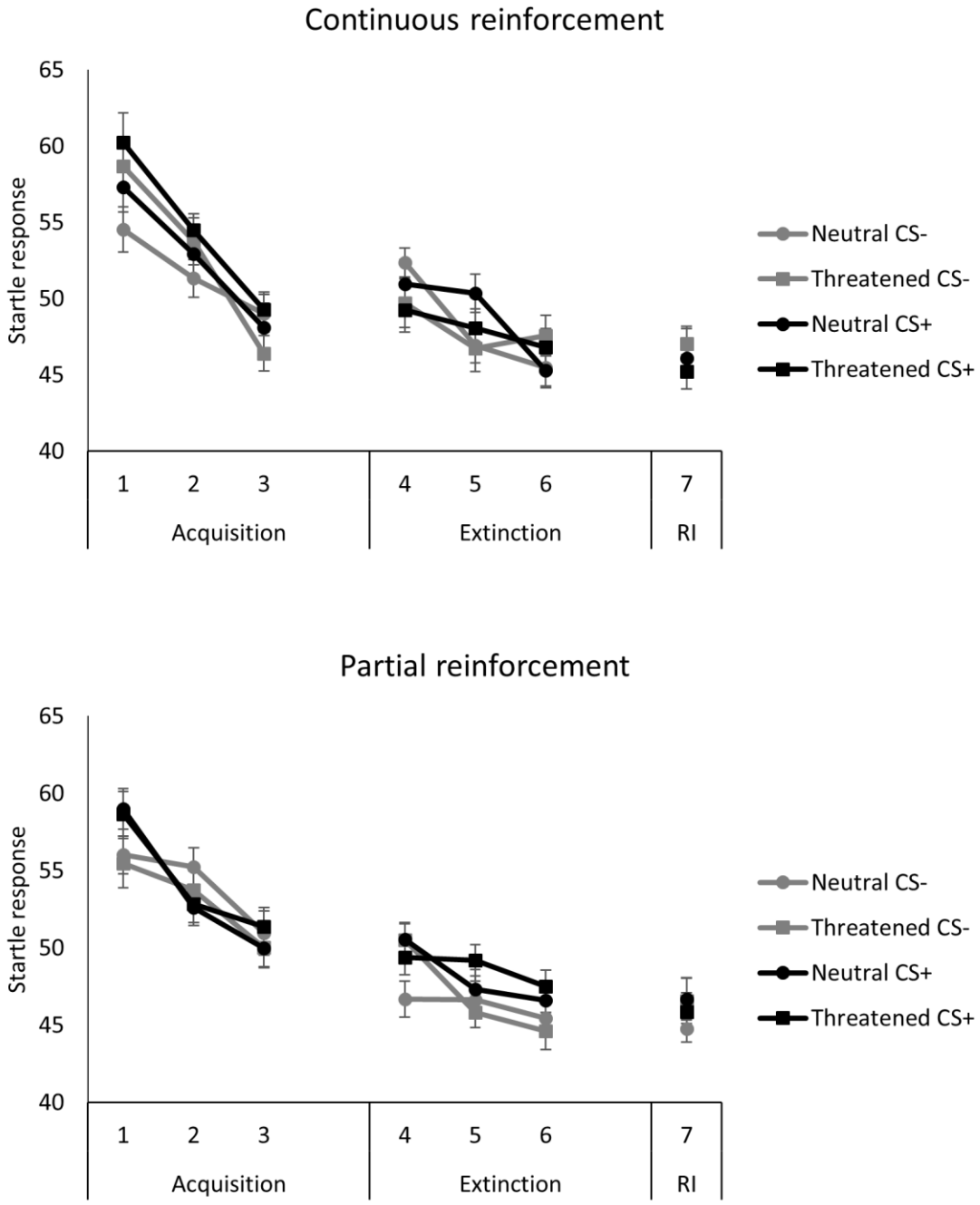

Figure 4. T-transformed startle responses (averaged per two trials) throughout the different phases of the experiment. RI = reinstatement.

Discussion

In the current study we investigated the impact of prior threatening information on subsequent fear acquisition, extinction and reinstatement. According to the expectancy bias model of Davey (1992, 1997) verbal threat can install a priori biases to expect aversive events (i.e., an aversive US) which can result in facilitated fear learning when USs are administered, and in delayed extinction of fear when USs are no longer applied. In addition to this model, we proposed that the effects of expectancy biases would be most pronounced in an uncertain situation because it has been demonstrated that cognitive biases (such as expectancy biases) exert their largest influence in such situations (Chan & Lovibond, 1996; Lissek et al., 2006; Ly & Roelofs, 2009; Tversky & Kahneman, 1974).

Some of the results of our study provide support for this interaction between verbal threatening instructions and stimulus pairings as predicted by Davey's (1992, 1997) expectancy model. First, verbal instructions installed a priori expectancy biases to expect the electric stimulus at the start of the conditioning phase. That is, higher US expectancy ratings on the first trial of the conditioning phase[7] were obtained for the threatened CSs compared to the control CSs, $F(1, 59) = 42.72$, $p < .001$, $\eta^2_p = .42$ (see Figure 2). Furthermore, these expectancy biases were also translated in higher psychophysiological fear responses at the start of the conditioning phase, as illustrated by the stronger startle reactions to the threatened CSs compared to the

[7] Note that on the first trial of the acquisition phase participants have no information about the contingency of the CSs with the US because the fear measures are collected during CS presentation, while reinforcement occurred at CS offset. Hence, US expectancy ratings on the first acquisition trial can be taken to reflect US expectancy "prior" to stimulus pairings. This is also illustrated by the absence of a main effect of CS type on the first trial of the acquisition phase, $F < 1$.

control CSs on the first trial of the conditioning phase, $F(1, 54) = 6.07$, $p = .017$, $\eta^2_p = .10$[8] (see Figure 4). Furthermore, significant interactions between threatening information and learning through stimulus pairings were observed in the extinction phase. First, conditioned fear reactions as measured with the startle reactions were more resistant to extinction for the threatened CS+ compared to the control CS+. That is, differential startle responses (i.e., higher for CS+ than for CS-) were still observed for the threatened CSs in middle and last part of the extinction phase in partial reinforcement condition (one-sided *p*-values: .003 and .060, respectively), but not for the neutral CSs (one-sided *p*-values: .440 and .203, respectively). Second, an expectancy bias was obtained for the threatened CS- compared to the neutral CS- at the start of the extinction phase. That is, US expectancies were higher for the threatened CS- than for the neutral CS- during the first two trials of the extinction phase in the partial reinforcement condition (one-sided *p*-values: .048 and .054, respectively). Finally, the fact that these biasing effects of verbal instructions were primarily observed in a more ambiguous situation (i.e., in the condition with the partial reinforcement schedule schedule), but not in a more unambiguous situation (i.e., in the condition with continuous reinforcement schedule), is in line with prior research about cognitive biases in general and expectancy biases in the context of fear more specific (Chan & Lovibond, 1996; Ly & Roelofs, 2009; Tversky & Kahneman, 1974).

Both from a clinical and theoretical perspective, these findings have interesting implications. Clinically, these results indicate verbal threatening instructions can amplify the effects of an aversive learning experience, resulting in more fear than when no threatening instructions were given. Thus, this implicates that care should be taken with providing

---

[8] This main effect of threat information remained significant when only the first trial was investigated rather than the average of the first two trials, $F(1, 48) = 5.74$, $p = .021$, $\eta^2_p = .11$.

(excessive) verbal warnings and threats because these might install expectancy biases and could amplify the effects of aversive experience. Indeed, research on anxiety-related disorders indicates that an over-protective parenting style, which usually involves excessive verbal warnings and threats, is a predictor of anxiety disorders such as social phobia (Rapee & Heimberg, 1997). Theoretically, our results indicate that biased fear learning can also occur when the threatening aspects of a stimulus are acquired on the basis of verbal instructions. Previously, biased fear learning for certain stimuli (such as pictures of spiders and snakes) has been attributed to the phylogenetic survival-relevance of these stimuli which allows these stimuli to be processed by a fast, efficient and unconscious fear learning system (Öhman & Mineka, 2001). However, Davey (1992, 1997) argued that such biased fear learning is in fact likely due to (conscious and reportable) expectancy biases and that these biases can also be acquired on the basis of cultural learning. The results of our experiment do indeed provide some evidence for this theory of Davey as we outlined before. Our results thus demonstrate that phylogenetic preparedness of stimuli is not a necessary condition for biased fear learning.

However, we have to note that the evidence that we provided in support of the expectancy bias model should be regarded as tentative for several reasons. First, the results in support of the expectancy bias model are inconsistent across the different measures and phases in this experiment. That is, the expected interaction between threat instructions and CS type was only observed for US expectancy ratings and FPS, but not for SCRs, and this interaction was only observed during the extinction phase, but not during the acquisition phase. The fragmented nature of the evidence suggests the possibility that the interactions we did observed are due to a type 1 error (false positive). We did not correct for multiple testing for the different phases and measures in our experiment because this would result in a very conservative alpha level and

would thus penalize the use of multiple measures of conditioned fear. Indeed, it is not common in fear conditioning research to control for multiple testing when conditioned fear is measured with different fear measures and in different phases. Nonetheless, care should be taken when interpreting the current results due to the possibility of type 1 errors. Second, sometimes the observed effects were in the opposite direction than what would be predicted on the basis of the expectancy bias model. That is, SCRs towards the neutral animals were slightly stronger than those towards the threatened animals just before and after the reinstatement manipulation ($p = .061$). A similar trend was observed in the condition with the continuous reinforcement schedule during the acquisition phase ($p = .085$). Although these results should be interpreted cautiously because they could also reflect false positives, they also call for caution when interpreting small but significant results in favor of the expectancy bias model. Finally, it is possible that the results could (partly) reflect experimental demand effects. That is, participants may infer on the basis of the instructions that they should use the threat instructions about the animals to indicate their expectancies about receiving an electric stimulus, without actually believing that the threat instructions are relevant for predicting the electric stimulus. Furthermore, the fact that participants had to repeatedly provide expectancy ratings about the electric stimulus while seeing the threatened and neutral animals probably encouraged participants even more to use the threat information to provide their expectancy ratings. Hence, at least for the biased US expectancy ratings to the threatened CSs it is possible that experimental demand operated. However, the finding that also the startle reflexes seemed to have been influenced by the threat instructions (on the first trial of the acquisition phase as well as during the extinction phase) provides some evidence against the interpretation that the obtained results were merely due to demand compliance.

We can thus conclude that our data provides some support for the expectancy bias model of Davey (1992, 1997), even though the of evidence should be regarded as tentative. We discuss four methodological and procedural considerations here that may impact the size of expectancy biases on fear conditioning. First, the believability of the expectancy bias manipulation should be considered. Approximately one fifth of the participants in this study reported to find the threat instructions not very or not at all believable. This result indicates that our threat instruction manipulation was not very convincing for (a part of) the participants, resulting in smaller effects of this manipulation and a reduction of the statistical power due to the exclusion of participants that did not find the information believable. One way to increase the believability of the threat instructions could be to include sources for the threat information (e.g., links to newspaper articles or Wikipedia entries; see Ugland et al., 2013 for a convincing demonstration). Second, threat information was provided right before the acquisition phase and the effect of the combination of threat information and stimulus pairings on extinction and reinstatement of fear was evaluated immediately after the acquisition phase. However, threat information and the combination of threat information with stimulus pairings may require consolidation in memory before exerting their full effect. That is, sleep and related memory consolidation processes can impact the recall of conditioned (Menz et al., 2013) and extinguished (Pace-Schott et al., 2009) fear memories. Although no prior studies have investigated this, it is possible that also the effects of threat information and the effects of the interaction between threat information and stimulus pairings also (partly) depend on sleep and memory consolidation processes. Related to this, memory organization principles may contribute to the interactive effects of threat information and stimulus pairings in the sense that it may be easier for participants to remember that a threatened animal was followed by an US (because this is more consistent information) than that

they can remember that a neutral animal was followed by an US (because this is more inconsistent information), especially when multiple days have passed since the acquisition phase (see Kleider, Pezdek, Goldinger, & Kirk, 2008 for a demonstration of this idea). Therefore, it would be interesting in future studies to investigate the effects of verbal threat information on acquisition, extinction and return of fear in a multi-day paradigm. Third, as we demonstrate in this study, the uncertainty of the learning situation should be taken into account. Previous studies have demonstrated that the uncertainty of the situation modulates the effect of expectancy biases (Chan & Lovibond, 1996; Lissek et al., 2006; Ly & Roelofs, 2009). In an unambiguous situation, there is little room for expectancy biases to modulate conditioned responses because the situation will dictate what appropriate responses are (Lissek et al., 2006). In contrast, in a more ambiguous situation there is more opportunity for the effects of expectancy biases to be expressed. Indeed, in our own study the effects of threat instructions were more outspoken in the condition using a partial reinforcement scheme for the electric stimulus, which produces a more ambiguous situation, than in the condition with a continuous reinforcement scheme. Thus, future studies focusing on the effects of expectancy biases in fear conditioning should preferable focus on more ambiguous situations, for instance by using a partial reinforcement scheme. Fourth and final, a between-subjects design may be more suited to find effects of expectancy biases on fear conditioning. Our study used a within-subjects design for manipulating threat information. While this design was preferred to increase statistical power, it may not be optimal for the purpose of the experiment because it provides participants with a benchmark to evaluate the effects of prior threat against. That is, because the same participants are exposed to both threatened and neutral CSs they can directly compare the contingency of the US with both these CSs, thereby quickly eroding any expectancy biases for the threatened CSs. Therefore, future studies may consider

employing a between-subjects design to manipulate threat information (Kahneman, 2011). Indeed, previous experiments focusing on differences in fear learning with fear-relevant and fear-irrelevant stimuli have often used between-subjects designs (Hugdahl & ?hman, 1977; Öhman, Erixon, & Lofberg, 1975), although more recently within-subjects designs have been used as well (Ho & Lipp, 2014; Lipp, Cronin, Alhadad, & Luck, 2015; Olsson, Ebert, Banaji, & Phelps, 2005). However, when a between-subjects design is used care should be taken to control for other possible factors that may vary between the conditions (such as elevated state anxiety due to threatening instructions) because these factors may introduce differences between the conditions other than the intended manipulation (see also Mertens, Raes, & De Houwer, 2016 for a discussion).

In summary, our study provides some tentative evidence that prior verbal threat information and subsequent stimulus pairings can interact. In line with the expectancy model of Davey (1992, 1997) our results demonstrate that verbal threat information can install expectancy biases that produce higher physiological fear reactions prior to stimulus pairings and delay extinction of fear. Our results further demonstrate that the effects of these expectancy biases were more pronounced in a more ambiguous learning situation (partial reinforcement) than in an unambiguous learning situation (continuous reinforcement). These results shed light on the conditions under which biased fear learning can take place.

Acknowledgements

The research reported in this paper was funded by the Interuniversity Attraction Poles Program initiated by the Belgian Science Policy Office (IUAPVII/33) and by Ghent University Methusalem Grant BOF09/01M00209.

References

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V, & van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, *42*(1), 1–15. https://doi.org/10.1111/j.1469-8986.2005.00271.x

Chan, C. K. Y., & Lovibond, P. F. (1996). Expectancy bias in trait anxiety. *Journal of Abnormal Psychology*, *105*(4), 637–647. https://doi.org/10.1037//0021-843X.105.4.637

Davey, G. C. L. (1992). An expectancy model of laboratory preparedness effects. *Journal of Experimental Psychology: General*, *121*(1), 24–40. https://doi.org/10.1037/0096-3445.121.1.24

Davey, G. C. L. (1997). A conditioning model of phobias. In *Phobias: a handbook of theory, research and treatment* (pp. 301–322).

Dawson, M. E., Schell, A. M., Filion, D. L., & Berntson, G. G. (2007). The Electrodermal System. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of Psychophysiology* (3rd ed., pp. 157–181). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511546396

De Clercq, A., Verschuere, B., De Vlieger, P., & Crombez, G. (2006). Psychophysiological Analysis (PSPHA): A modular script-based program for analyzing psychophysiological data. *Behavior Research Methods*, *38*(3), 504–510. https://doi.org/10.3758/BF03192805

Field, A. P. (2006). Is conditioning a useful framework for understanding the development and treatment of phobias? *Clinical Psychology Review*, *26*(7), 857–875. https://doi.org/10.1016/j.cpr.2005.05.010

Field, A. P., & Storksen-Coulson, H. (2007). The interaction of pathways to fear in childhood anxiety: a preliminary study. *Behaviour Research and Therapy*, *45*(12), 3051–3059. https://doi.org/10.1016/j.brat.2007.09.001

Hamm, A. O., & Weike, A. I. (2005). The neuropsychology of fear learning and fear regulation. *International Journal of Psychophysiology*, *57*(1), 5–14. https://doi.org/10.1016/j.ijpsycho.2005.01.006

Ho, Y., & Lipp, O. V. (2014). Faster acquisition of conditioned fear to fear-relevant than to nonfear-relevant conditional stimuli. *Psychophysiology*, *51*(8), 810–813. https://doi.org/10.1111/psyp.12223

Hugdahl, K., & ?hman, A. (1977). Effects of instruction on acquisition and extinction of electrodermal responses to fear-relevant stimuli. *Journal of Experimental Psychology: Human Learning & Memory*, *3*(5), 608–618. https://doi.org/10.1037/0278-7393.3.5.608

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

King, N. J., Eleonora, G., & Ollendick, T. H. (1998). Etiology of childhood phobias: Current status of Rachman's three pathways theory. *Behaviour Research and Therapy*, *36*(3), 297–309. https://doi.org/10.1016/S0005-7967(98)00015-1

Kleider, H. M., Pezdek, K., Goldinger, S. D., & Kirk, A. (2008). Schema-driven source misattribution errors: remembering the expected from a witnessed event. *Applied Cognitive Psychology*, *22*(1), 1–20. https://doi.org/10.1002/acp.1361

Kuhn, M., Mertens, G., & Lonsdorf, T. B. (2016). State anxiety modulates the return of fear. *International Journal of Psychophysiology*, *110*, 194–199. https://doi.org/10.1016/j.ijpsycho.2016.08.001

Lang, P. J. (1968). Fear reduction and fear behavior: Problems in treating a construct. In J. M. Schlien (Ed.), *Research in psychotherapy* (pp. 90–103). Washington, DC: American Psychological Association.

Lipp, O. V., Cronin, S. L., Alhadad, S. S. J., & Luck, C. C. (2015). Enhanced sensitization to animal, interpersonal, and intergroup fear-relevant stimuli (but no evidence for selective one-trial fear learning). *Psychophysiology*, *52*(11), 1520–1528. https://doi.org/10.1111/psyp.12513

Lissek, S., Pine, D. S., & Grillon, C. (2006). The strong situation: A potential impediment to studying the psychobiology and pharmacology of anxiety disorders. *Biological Psychology*, *72*(3), 265–270. https://doi.org/10.1016/j.biopsycho.2005.11.004

Ly, V., & Roelofs, K. (2009). Social anxiety and cognitive expectancy of aversive outcome in avoidance conditioning. *Behaviour Research and Therapy*, *47*(10), 840–7. https://doi.org/10.1016/j.brat.2009.06.015

Lykken, D. T., & Venables, P. H. (1971). Direct measurement of skin conductance: a proposal for standardization. *Psychophysiology*, *8*(5), 656–672. https://doi.org/10.1111/j.1469-8986.1971.tb00501.x

McNally, R. J. (1987). Preparedness and phobias: A review. *Psychological Bulletin*, *101*(2), 283–303. https://doi.org/10.1037/0033-2909.101.2.283

Menz, M. M., Rihm, J. S., Salari, N., Born, J., Kalisch, R., Pape, H. C., … Büchel, C. (2013). The role of sleep and sleep deprivation in consolidating fear memories. *NeuroImage*, *75*, 87–96. https://doi.org/10.1016/j.neuroimage.2013.03.001

Merckelbach, H., de Jong, P. J., Muris, P., & van den Hout, M. A. (1996). The etiology of

specific phobias: A review. *Clinical Psychology Review*, *16*(4), 337–361. https://doi.org/10.1016/0272-7358(96)00014-1

Mertens, G., & De Houwer, J. (2016). Potentiation of the startle reflex is in line with contingency reversal instructions rather than the conditioning history. *Biological Psychology*, *113*, 91–99. https://doi.org/10.1016/j.biopsycho.2015.11.014

Mertens, G., Raes, A. K., & De Houwer, J. (2016). Can prepared fear conditioning result from verbal instructions? *Learning and Motivation*, *53*, 7–23. https://doi.org/10.1016/j.lmot.2015.11.001

Milad, M. R., Orr, S. P., Pitman, R. K., & Rauch, S. L. (2005). Context modulation of memory for fear extinction in humans. *Psychophysiology*, *42*(4), 456–64. https://doi.org/10.1111/j.1469-8986.2005.00302.x

Mineka, S., & Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorders: it's not what you thought it was. *The American Psychologist*, *61*(1), 10–26. https://doi.org/10.1037/0003-066X.61.1.10

Morriss, J., Christakou, A., & van Reekum, C. M. (2016). Nothing is safe: Intolerance of uncertainty is associated with compromised fear extinction learning. *Biological Psychology*, *121*, 187–193. https://doi.org/10.1016/j.biopsycho.2016.05.001

Muris, P., & Field, A. P. (2010). The role of verbal threat information in the development of childhood fear. "Beware the Jabberwock!" *Clinical Child and Family Psychology Review*, *13*(2), 129–150. https://doi.org/10.1007/s10567-010-0064-1

Nitschke, J. B., Miller, G. a., & Cook, E. W. (1998). Digital filtering in EEG/ERP analysis: Some technical and empirical comparisons. *Behavior Research Methods, Instruments, &*

Computers, *30*(1), 54–67. https://doi.org/10.3758/BF03209416

Öhman, A., Erixon, G., & Lofberg, I. (1975). Phobias and preparedness: Phobic versus neutral pictures as conditioned stimuli for human autonomic responses. *Journal of Abnormal Psychology*, *84*(1), 41–45. https://doi.org/10.1037/h0076255

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522. https://doi.org/10.1037//0033-295X.108.3.483

Olsson, A., Ebert, J. P., Banaji, M. R., & Phelps, E. a. (2005). The role of social groups in the persistence of learned fear. *Science (New York, N.Y.)*, *309*(5735), 785–7. https://doi.org/10.1126/science.1113551

Olsson, A., & Phelps, E. A. (2004). Learned fear of "unseen" faces after Pavlovian, observational, and instructed fear. *Psychological Science*, *15*(12), 822–828. https://doi.org/10.1111/j.0956-7976.2004.00762.x

Pace-Schott, E. F., Milad, M. R., Orr, S. P., Rauch, S. L., Stickgold, R., & Pitman, R. K. (2009). Sleep promotes generalization of extinction of conditioned fear. *Sleep*, *32*(1), 19–26.

Pineles, S. L., Orr, M. R., & Orr, S. P. (2009). An alternative scoring method for skin conductance responding in a differential fear conditioning paradigm with a long-duration conditioned stimulus. *Psychophysiology*, *46*(5), 984–95. https://doi.org/10.1111/j.1469-8986.2009.00852.x

Rachman, S. (1977). The conditioning theory of fear-acquisition: a critical examination. *Behaviour Research and Therapy*, *15*(5), 375–387. https://doi.org/10.1016/0005-7967(77)90041-9

Raes, A. K., De Houwer, J., De Schryver, M., Brass, M., & Kalisch, R. (2014). Do CS-US Pairings Actually Matter? A Within-Subject Comparison of Instructed Fear Conditioning with and without Actual CS-US Pairings. *PloS One*, *9*(1), e84888. https://doi.org/10.1371/journal.pone.0084888

Rapee, R. M., & Heimberg, R. G. (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*, *35*(8), 741–756. https://doi.org/10.1016/S0005-7967(97)00022-3

Schindler, B., Vriends, N., Margraf, J., & Stieglitz, R.-D. (2016). Ways of acquiring flying phobia. *Depression and Anxiety*, *33*(2), 136–142. https://doi.org/10.1002/da.22447

Seligman, M. E. P. (1971). Phobias and preparedness. *Behavior Therapy*, *2*(3), 307–320. https://doi.org/10.1016/S0005-7894(71)80064-3

Soeter, M., & Kindt, M. (2012). Erasing fear for an imagined threat event. *Psychoneuroendocrinology*, *37*(11), 1769–1779. https://doi.org/10.1016/j.psyneuen.2012.03.011

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*(3), 495. https://doi.org/10.1017/S0140525X0003123X

Torrents-Rodas, D., Fullana, M. a, Bonillo, A., Caseras, X., Andión, O., & Torrubia, R. (2013). No effect of trait anxiety on differential fear conditioning or fear generalization. *Biological Psychology*, *92*(2), 185–90. https://doi.org/10.1016/j.biopsycho.2012.10.006

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Ugland, C. C. O., Dyson, B. J., & Field, A. P. (2013). An ERP study of the interaction between verbal information and conditioning pathways to fear. *Biological Psychology*, *92*(1), 69–81. https://doi.org/10.1016/j.biopsycho.2012.02.003

van der Ploeg, H. M., Defares, P. B., & Spielberger, C. D. (2000). *Handleiding bij de Zelfbeoordelings Vragenlijst. Een Nederlandstalige bewerking van de Spielberger State-Trait Anxiety Inventory*. Lisse, The Netherlands.

Appendix

English translation of the Dutch threat information:

Threat information: The two animals pictured above are the Pizote and Cuscus. These two animals live in Australia and are marsupials. The Pizote and Cuscus are known as aggressive animals. They often go to the suburbs of cities to find food between the trash there and will, while doing so, sometimes attack pets and even humans. The Australian government advices to keep a safe distance from the Pizote and Cuscus because they are known to carry diseases such as rabies and the Marburg virus.

Neutral information: The two animals pictured above are the Ayeaye and the Quoll. These two animals live in the densely forested areas of Middle- and South-America. Because the habitat of these animals is so remote, they only rarely come into contact with humans. The diet of these animals mainly consists of nuts, roots and tubers. The Ayeaye and Quoll are both nocturnal animals and mainly go searching for food at dusk or during the night.

Supplementary Material

Most of the correlations between the acquisition, extinction and reinstatement indexes were only weak and non-significant, reflecting similar inconsistent findings in the literature (Kuhn et al., 2016; Morriss, Christakou, & van Reekum, 2016; Torrents-Rodas et al., 2013). A few of the correlations did reach the uncorrected significance threshold. Particularly, significant correlations were obtained between STAI-T scores and US expectancy ratings, indicating less extinction of the neutral CS and less reinstatement of the neutral CS in the condition with a partial reinforcement schedule. Furthermore, correlations between the STAI-T scores and fear ratings indicated more acquisition for the neutral CS+ in the condition with the continuous reinforcement schedule. Finally, correlations between the STAI-T scores and SCRs indicated more acquisition for the neutral CS+ in the condition with the partial reinforcement schedule and more reinstatement for the threatened CS- in the condition with the continuous reinforcement schedule. No significant correlations between STAI-T scores and FPS acquisition, extinction or reinstatement indexes were observed. Importantly, the fact that the observed significant correlations were not consistently observed for a particular type of CS, in a particular phase or with either a continuous or partial reinforcement schedule indicates that great care should be taken when interpreting these results and that they likely reflect false positives. These equivocal findings observed here, as well as those observed in earlier studies, may indicate that the STAI-T does not essentially measures the cognitive mechanisms involved in (maladaptive) fear acquisition, extinction or reinstatement (see Morriss et al., 2016 for a discussion).

Table 1

Correlations between mean centered STAI-T scores and US expectancy rating indexes. $^{+}$ *= p < .10; * = p < .05; ** = p < .01*

	Continuous reinforcement	Partial reinforcement	Difference (*F*-value)	Interaction threat*reinforcement
Acquisition neutral CS	-.097	.136	< 1	
Acquisition threatened CS	-.073	.210	1.40	
Difference (*F*-value)	< 1	< 1		< 1
Extinction neutral CS	-.051	.443*	5.86*	
Extinction threatened CS	-.078	.272	2.21	
Difference (*F*-value)	< 1	2.62		1.97
Reinstatement neutral CS+	.036	-.411*	4.65*	
Reinstatement threatened CS+	.231	-.151	2.09	
Difference (*F*-value)	1.47	1.88		< 1
Reinstatement neutral CS-	.096	.224	< 1	
Reinstatement threatened CS-	-.155	-.204	< 1	
Difference (*F*-value)	1.65	2.39		< 1

Table 2

Correlations between mean centered STAI-T scores and fear rating indexes. $^{+}$ *= p < .10; * = p < .05; ** = p < .01*

	Continuous reinforcement	Partial reinforcement	Difference (*F*-value)	Interaction threat*reinforcement
Acquisition neutral CS	.436*	.120	< 1	
Acquisition threatened CS	.187	.078	< 1	
Difference (*F*-value)	< 1	< 1		< 1

Table 3

Correlations between mean centered STAI-T scores and SCR indexes. $^{+}$ *= p < .10; * = p < .05; ** = p < .01*

	Continuous reinforcement	Partial reinforcement	Difference (*F*-value)	Interaction threat*reinforcement
Acquisition neutral CS	.307	.396*	< 1	
Acquisition threatened CS	-.092	-.153	< 1	
Difference (*F*-value)	2.27	5.82*		< 1
Extinction neutral CS	-.239	.090	< 1	
Extinction threatened CS	.214	-.075	< 1	
Difference (*F*-value)	3.60^{+}	< 1		2.74
Reinstatement neutral CS+	-.170	.132	< 1	
Reinstatement threatened CS+	.006	-.055	< 1	
Difference (*F*-value)	< 1	< 1		< 1
Reinstatement neutral CS-	-.346	.081	1.79	
Reinstatement threatened CS-	.588*	.201	7.12*	
Difference (*F*-value)	13.62**	1.10		8.80**

Table 4

Correlations between mean centered STAI-T scores and FPS indexes. $^{+}$ *= p < .10; * = p < .05; ** = p < .01*

	Continuous reinforcement	Partial reinforcement	Difference (*F*-value)	Interaction threat*reinforcement
Acquisition neutral CS	.237	.195	< 1	
Acquisition threatened CS	-.235	.048	< 1	
Difference (*F*-value)	2.91	< 1		< 1
Extinction neutral CS	-.137	.044	< 1	
Extinction threatened CS	-.134	.352	4.39*	
Difference (*F*-value)	< 1	1.44		1.01
Reinstatement neutral CS+	-.090	.093	< 1	
Reinstatement threatened CS+	-.019	.103	< 1	
Difference (*F*-value)	< 1	< 1		< 1
Reinstatement neutral CS-	-.068	.112	< 1	
Reinstatement threatened CS-	-.251	.044	< 1	
Difference (*F*-value)	< 1	< 1		< 1